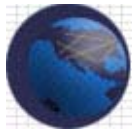


## Integrative Cancer Research Special Interest Group Teleconference

### Gene Annotation SIG Meeting Minutes

<b>Date, Time &amp; Location:</b>	May 6, 2004 1:00 – 2:00 EDT
<b>Attendees:</b>	<p>Patrick McConnell – Duke (funded developer)  Simon Lin – Duke (funded developer)  Kutbuddin Doctor – Burnham (funded developer)  Terry Braun – Holden (funded developer)  Cathy Wu – Georgetown (funded developer)  Terry Disz – U of C (funded developer)  Rakesh Nagarajan – Wash U (funded developer)  Don Baldwin – Penn (funded adopter)  David Fenstermacher – Penn (funded adopter)  Harold Riethman – Wistar (funded adopter)  David Kane – SRA/CCR  Phan Winer – BAH  Juli Klemm – BAH</p> <p><u>Computational Genomics attendees:</u>  Patrick McConnell – Duke (funded developer)  Simon Lin – Duke (funded developer)  Harold Riethman – Wistar (funded adopter)  Vincent Yau – Oregon Health (funded adopter)  Claire Zhu – BAH  Juli Klemm - BAH</p>
<b>Introduction:</b>	<p><b><u>Roll-call, open meeting, review meeting goals</u></b></p> <ul style="list-style-type: none"> <li>- Establish goals and priorities for this SIG</li> <li>- For match making purposes, Developers and Adopters will be asked to give a brief statement of their capabilities and interests, respectively</li> <li>- Identify and define additional research</li> </ul>
<b>Overview Discussion:</b>	<p><b><u>Review goals and objectives of Gene Annotation SIG</u></b></p> <p>Scope: Tools and data sources for Gene Annotation</p> <p><b><u>Open Discussion</u></b></p> <ul style="list-style-type: none"> <li>- “Gene annotation” is a vague term. Most broadly speaking it can be extended to include not only functional annotation but also the locations of SNPs within genes as well as haplotype information.</li> <li>- The caBIO objects are a rich source of annotation information but this would not be the only way to share annotation information within caBIG. Having agreed-upon interfaces for sharing data is the most important consideration.</li> <li>- A key issue with regard to gene annotation is the problem of unstable gene identifiers. <ul style="list-style-type: none"> <li>o One example solution: the UniProt consortium stores all gene objects in a universal archive. Users can track the history of all records that were ever related to a gene/protein.</li> </ul> </li> <li>- Gene naming conventions like HUGO are nice in theory, but one must still deal</li> </ul>



## Integrative Cancer Research Special Interest Group Teleconference

with the real issue of synonyms. Relatively stable identifiers include:

- Genbank accession
- LocusID (very concrete to link to a genome location)
- Ensemble ID
- It was recommended that it would be useful if web services were available to link between RefSeq IDs, LocusLink IDs and HUGO.
  - PIR provides this type of service
- Regardless of the most reliable mapping method, there is still often ambiguity that needs to be solved. Some suggestions:
  - Provide some sort of confidence measure to the user
  - Provide all available information to the user and let them determine how to interpret
    - Answer might depend on the context of the question being asked.
- The group agreed that GO annotation is extremely useful to their user community
  - Essential to have GO computation on pathways
  - Useful to have GO annotation for microarray data meta-analysis
  - Users should be cautious – there is lack of uniformity within GO
    - It is important to capture the evidence code with the GO annotation.

### High-level review of Center interests in this SIG

#### Developers

U of C: SEED is a self-contained annotation system. Allows individual scientists to create their own annotations. It is collaborative in the sense that any SEED user can connect to any other SEED user in a peer-to-peer network. It is currently focused on prokaryotes – the caBIG proposal is to move it to eukaryotes.

Holden: TrAPSS is a sequence – based annotation tool. Used to prioritize therapeutic candidates. Pulls data from a variety of sources; performs a variety of analyses (e.g. transmembrane prediction).

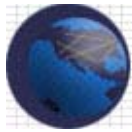
Georgetown: PIR is a mature database containing a great deal of actively-curated protein annotation. Uses a protein-family classification approach. Interactive interface coupled with an annotation pipeline. For caBIG, will develop a large-scale proteomics and genomics interface for adopters to find comprehensive data.

Wash U: Annotation information available from a variety of sources – using a standard identifier, can perform high-throughput annotation of microarray data. Annotations cross organisms to allow for a very broad level of annotation. Also have available a tool to discover literature (abstracts) co-occurrences of genes and to create a network based on this information. Wash U also has available a mutational profile tool.

Burnham: Molecular Pages system allows users to look at gene information in the context of many experiments. The technology currently supports the JCSG project – would make cancer-specific for this project. See example pages at:

[http://www.jcsg.org/scripts/prod/public\\_targets/pub\\_target\\_list.cgi?](http://www.jcsg.org/scripts/prod/public_targets/pub_target_list.cgi?)

[show\\_pdb=checked&show\\_human=checked&show\\_model=checked&show\\_desc=checked](http://www.jcsg.org/scripts/prod/public_targets/pub_target_list.cgi?show_pdb=checked&show_human=checked&show_model=checked&show_desc=checked)



## Integrative Cancer Research Special Interest Group Teleconference

Center for Cancer Research: MatchMiner: Allows user to translate between disparate IDs for the same gene. Has a web-based UI as well as a command-line UI. With regard to the discrepancy question, the system flags discrepancies and tracks which database each identifier comes from to let the user track down how it should be resolved. GOMiner; A “consumer” of annotations. Analyzes GO assignments of a given list of genes, generally in the context of some change.

Cold Spring: The International Human HapMap Project seeks to map out regions of common genetic variability in the human genome by genotyping three major world populations at a resolution of 1 marker per 5 kbp. The Stein Lab manages the central database for the project, allocates SNPs to the 11 genotyping centers, coordinates data submission, performs quality checks and quality control, and manages the public release of project data. Access to the data is provided both for bulk download and for interactive querying and browsing using the GMOD tools.

From Computational Genomics Meeting:

### Duke:

- JProbePicker: an application for choosing oligo probes for custom microarrays. Open architecture. Allows plug-ins of other filters/algorithms for probe ranking. Coded in Java, and driven by XML.
- JProgram: Java programming library that would allow caBIG developers to describe their tools. Open source version is currently available on their website. Working on APIs that provide access to existing tools.
- JOrganBlast: Developed as a proof-of-concept for the caBIG site visit: webservices linked in a workflow – output of one tool piped in as input to another. Filters gene expression data by organ. Short-term goal is to make this tool caBIG compatible. Drag/Drop type of desktop services as a long term goal.

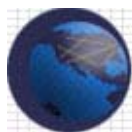
### Adopters

Duke: Is performing a large number of microarray studies. Generally get the analysis down to a list of 50 – 60 genes, then would like to understand them in more depth. In particular, would like to know what other genes could be cross-hybridizing to contribute to the signal. [Wash U stated they have been thinking about this problem as well and have worked on a method to comprehensively ask this question – very computationally intensive.]

Penn: Many interests: Transcript profiling, proteomics, genotyping.

- Would like to use public repositories to more efficiently query against public transcript profiles.
- Currently using Ingenuity. Would like to see if this company would release its tool to be used more broadly.
- Likes EASE functionality. This is a free tool; would like to see it integrated with other packages
- PIR tools seem very interesting
- Generally interested in tools that can be altered quickly such that a new data source can be brought in.

Wistar: Very interested in HAPMAP-type data. A centralized source for SNP data would be very useful; important for association studies. Also interested in Molecular Pages, GOMiner, MatchMiner.



## Integrative Cancer Research Special Interest Group Teleconference

### Other discussion items:

#### Mission Statement

Kutbuddin Doctor from Burnham will draft a Mission Statement for this SIG. This draft will be circulated to the group for comments.

#### Regular meeting schedule

The group will meet on the first Thursday of each month at 2:00 Eastern. Ongoing frequency will be revisited if necessary.

#### Notes and additional information

Meeting minutes will be emailed and posted on the caBIG on-line forum after review by the group:

[http://ncicbforums.nci.nih.gov/forums/cabigforum/lfs/icrlfs/SIGs/gene\\_annotation](http://ncicbforums.nci.nih.gov/forums/cabigforum/lfs/icrlfs/SIGs/gene_annotation)

Developers should also post additional information about their tools here.

### Action Items:

Name Responsible	Action Item	Date Due	Notes
Kutbuddin Doctor	Draft Gene Annotation SIG mission statement	5/28/04 (approx)	
Juli Klemm	Distribute meeting minutes	5/14/04	
Juli Klemm	Schedule ongoing meetings	5/14/04	
All developers	Post relevant additional information to the caBIG on-line forum (see link above)	5/14/04	
Juli Klemm	Arrange for Wash U Mutation Profiler presentation to Wistar and Penn	5/14/04	